

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-240739

(43)Date of publication of application : 11.09.1998

(51)Int.Cl.

G06F 17/30

(21)Application number : 09-043577

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 27.02.1997

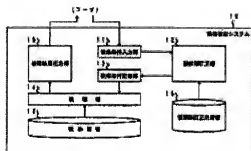
(72)Inventor : SAKAI TETSUYA
SUMITA KAZUO

(54) DEVICE FOR RETRIEVING INFORMATION AND METHOD THEREFOR

(57)Abstract:

PROBLEM TO BE SOLVED: To provide an information retrieving device in which retrieval can be executed by automatically detecting and correcting an erroneously inputted keyword phrase.

SOLUTION: A dictionary 16 for retrieval word correction is a dictionary including synonym knowledge for obtaining the synonym of a phrase inputted as a keyword, co-occurrence word knowledge for obtaining a co-occurrence word, and related term knowledge for obtaining a related term or the like. A retrieval word correcting part 12 detects an erroneously inputted phrase included in a retrieval condition inputted through a retrieval condition inputting part 11 by referring to the dictionary 16 for retrieval word correction. Then, the retrieval word correcting part 12 corrects this detected erroneously inputted phrase by the dictionary 16 for retrieval word correction, and stores the retrieval condition including the corrected phrase in a retrieval condition storing part 13. Thus, a retrieving part 14 is allowed to execute retrieval using the corrected phrase.



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-240739

(43) 公開日 平成10年(1998) 9月11日

(51) Int.Cl.⁸

識別記号

F I

G 0 6 F 17/30

G 0 6 F 15/403

3 2 0 D

審査請求 未請求 請求項の数 5 O L (全 11 頁)

(21) 出願番号 特願平9-43577

(22) 出願日 平成9年(1997) 2月27日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 酒井 哲也

神奈川県川崎市幸区小向東芝町1番地 株

式会社東芝研究開発センター内

(72) 発明者 住田 一男

神奈川県川崎市幸区小向東芝町1番地 株

式会社東芝研究開発センター内

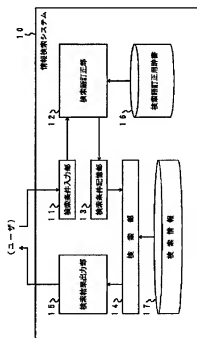
(74) 代理人 弁護士 鈴江 武彦 (外6名)

(54) 【発明の名称】 情報検索装置および情報検索方法

(57) 【要約】

【課題】 誤入力されたキーワード語句を自動的に検出・訂正して検索を実行する情報検索装置を提供する。

【解決手段】 検索語訂正用辞書16は、キーワードとして入力された語句の類似語を得るための類似語知識、共起語を得るための共起語知識および関連語を得るための関連語知識などを含む辞書であり、検索語訂正部12は、この検索語訂正用辞書16を参照することにより、検索条件入力部11を介して入力された検索条件に含まれる誤入力語句を検出する。そして、検索語訂正部12は、この検出した誤入力語句を検索語訂正用辞書16により訂正し、その訂正後の語句を含んで構成される検索条件を検索条件記憶部13に格納することによって、検索部14に訂正後の語句を用いた検索を実行させる。



【特許請求の範囲】

【請求項1】 キーワードとして入力された語句を含んで構成される検索条件に合致した情報を検索する情報検索装置において、

類似語知識、共起語知識および関連語知識の少なくとも一つ以上を含む語句訂正用辞書と、

誤入力された語句を前記検索語訂正用辞書により検出する誤入力語句検出手段と、

前記誤入力語句検出手段が検出した語句を前記検索語訂正用辞書により訂正する語句訂正手段とを具備してなることを特徴とする情報検索装置。

【請求項2】 キーワードとして入力された語句を含んで構成される検索条件に合致した情報を検索する情報検索装置において、

類似語知識、共起語知識および関連語知識の少なくとも一つ以上を含む語句選定用辞書と、

前記入力された語句が多義に解釈される語句であるときに、前記語句選定用辞書によりいずれかの解釈に選定する語句選定手段とを具備してなることを特徴とする情報検索装置。

【請求項3】 検索された情報それぞれの有用度を示す評価を入力する評価入力手段と、

前記評価入力手段により入力された評価を分析して前記語句を訂正するフィードバック手段とをさらに具備してなることを特徴とする請求項1または2記載の情報検索装置。

【請求項4】 類似語知識、共起語知識および関連語知識の少なくとも一つ以上を含む語句訂正用辞書を備え、

キーワードとして入力された語句を含んで構成される検索条件に合致した情報を検索する情報検索方法において、

誤入力された語句を前記検索語訂正用辞書により検出し、

この検出された語句を前記語句訂正用辞書により訂正し、

この訂正後の語句を含んで構築される検索条件を用いて検索を実行することを特徴とする情報検索方法。

【請求項5】 類似語知識、共起語知識および関連語知識の少なくとも一つ以上を含む語句選定用辞書を備え、

キーワードとして入力された語句を含んで構成される検索条件に合致した情報を検索する情報検索方法において、

前記入力された語句が多義に解釈される語句であるときに、前記語句選定用辞書によりいずれかの解釈に選定し、

この選定結果を前記検索条件に反映させた後に検索を実行することを特徴とする情報検索方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 この発明は、キーワードとし

て入力された語句を含んで構成される検索条件に合致した情報を検索する情報検索装置および情報検索方法に係り、特に誤入力された語句を自動的に検出・訂正して検索を実行する情報検索装置および情報検索方法に関する。

【0002】

【従来の技術】 近年、パーソナルコンピュータ、インターネット、電子図書館などの普及に伴い、個人がアクセスできる情報の量は増加する一方であり、また、情報のメディアもテキスト、画像、音声など多様化している。そして、このような状況により膨大な情報の中から求める情報のみを検索する高度な情報検索システムの需要が高まりつつある。

【0003】 通常の情報検索システムを利用する際、ユーザは、自分の要求をシステムに伝えるために、一つ以上の検索語を入力してシステムが理解できる形式の検索条件を構成する。システムは、「ユーザの要求＝検索条件」という仮定のもとで、検索対象の中から検索条件を満たす情報のみを選出して出力する。

【0004】 しかし、実際は「ユーザの要求＝検索条件」という仮定が成り立たないケースが多い。特に、キーボードや文字認識装置、音声認識装置などの入力装置を介して検索条件を指定する際に、検索語の誤入力のためにユーザの意図とはかけ離れた検索条件が作成されてしまうことがある。たとえば、かな漢字変換システムを用いて漢字の検索語を指定しようとしたときに、変換誤りによってまったく関係のない同音異義語を入力してしまう場合がある。

【0005】 有り得ないつづりの単語などに対しては、すべての検索語に対して単語辞引きを行なうことによって誤入力の検出・訂正が可能であるが、前述した同音異義語の例のように、語入力の結果、単語としては正しい語句が得られてしまった場合には、従来の情報検索システムでは、これを検出・訂正することはできなかった。このため、誤った検索条件により誤った検索結果が得られ、そこで初めてユーザが自分の誤入力に気づくといったケースが生じていた。さらに、検索語の誤入力によって誤った検索結果が得られても、なお、ユーザが自分で誤入力の誤りに気づかないケースもあった。

【0006】

【発明が解決しようとする課題】 このように、従来の情報検索システムでは、単語としては正しい語句が検索語として誤入力されてしまった場合に、これを検出・訂正することができないといった問題があった。

【0007】 この発明はこのような実情に鑑みてなされたものであり、誤入力された語句を自動的に検出・訂正して検索を実行する情報検索装置および情報検索方法を提供することを目的とする。

【0008】

【課題を解決するための手段】 この発明は、前述した目

的を達成するために、たとえばキーワードとして入力された語句の類似語を得るための類似語知識、共起語を得るための共起語知識および関連語を得るための関連語知識などを含む辞書を利用して、誤入力された語句を検出・訂正してから検索を実行するようにしたものである。

【0009】また、この辞書を利用して、多義に解釈された語句の解釈を選定することにより、検索の効率および精度を向上させるものである。さらに、ユーザによる検索結果の評価を分析し、この分析結果を誤入力された語句の検出・訂正にフィードバックすることにより、再検索の精度を向上させるものである。

【0010】このように、誤入力があった場合であっても、ユーザが本来意図した検索条件を作成して検索を実行することにより、適切な情報検索環境を提供することを可能とする。

【0011】

【発明の実施の形態】以下、図面を参照してこの発明の実施の形態を説明する。

(第1実施形態) まず、この発明の第1実施形態について説明する。図1に、第1実施形態に係る情報検索システムの構成を示す。図1に示したように、この情報検索システム10は、検索条件入力部11、検索語訂正部12、検索条件記憶部13、検索部14および検索結果出力部15からなる。ここで、検索条件入力部11は、キーボード、文字認識装置、音声認識装置などの入力装置に、検索結果出力部15は、ディスプレイ、プリンタなどの出力装置に、検索条件記憶部13は、主記憶やハードディスク装置などにそれぞれ対応し、検索部14および検索語訂正部12は、CPUによって実行制御されるプログラムに対応する。

【0012】ユーザにより検索条件入力部11に入力された検索条件は、検索語訂正部12に渡され、必要に応じてキーワードとなる語句、すなわち検索語の訂正が行なわれる。訂正された検索条件は、検索条件記憶部13に記憶され、この検索条件により検索部14が情報の検索を行なう。検索結果は、検索結果出力部15によりユーザに出力される。

【0013】従来の情報検索システムとこの情報検索システム10との違いは、前者では検索条件入力部11に入力された検索条件が直接検索条件記憶部13に渡されるのに対し、後者では検索条件が一旦検索語訂正部12に渡され、その後検索条件記憶部13に渡される点のみである。よって、検索語訂正部12の処理以外は既存のどのような検索システムでも構わない。ユーザが一つ以上の検索語を指定することにより検索を行なうものであれば、検索条件の指定方法、検索対象、検索方法などは問わない。さらに、検索語は日本語、英語の他、どのような言語でも構わない。以下、検索語訂正部12の動作を中心に説明する。

【0014】図2に、第1実施形態の検索語訂正部12

の処理の流れの一例を示す。検索語訂正部12は、まず検索条件入力部11からユーザの入力した検索条件を受取り(ステップA1)、これを解析して検索語を同定する(ステップA2)。そして、各検索語について以下の処理を行なう。

【0015】現在着目している検索語の類似語知識およびそれ以外の検索語の共起語知識と関連語知識を検索語訂正用辞書から取出す(ステップA4)。次に、これらの知識を用いて、現在着目している検索語が誤入力であるかどうかを判定する(ステップA5)。誤入力であると判定した場合は(ステップA5のY)、検索語を前述の類似語で置換えることにより検索条件を訂正する(ステップA6)。

【0016】なお、検索語の訂正を完全自動で行なうかわりに、誤入力の検出のみをまず行ない、ユーザに「訂正しますか?」などのメッセージを表示して、対話的に訂正を行なうようにしてもよい。

【0017】図3に、第1実施形態の検索語訂正用辞書16に登録されている類似語知識、共起語知識および関連語知識の一例を示す。この第1実施形態では、「語A」がかな漢字変換誤りやつづり誤りなどの誤入力によって「語B」になってしまうような場合に、「語A」と「語B」とは互いに類似語であるということにする。

【0018】図3の例では、類似語知識として、同音異義語知識と類似表記語知識との2種類が示されている。

図3の(知識a)〜(知識d)が同音異義語知識の例であり、(知識e)〜(知識g)が類似表記語知識の例である。たとえば、(知識a)から、「金庫」と「解析」という語がともに「かいせき」という読みをもつことがわかり、また、(知識g)から、「leader」と「reader」という語は表記が似ており誤入力しやすいということがわかる。この同音異義語知識は、たとえば既存のかな漢字変換用辞書を利用して構築することが可能である。また、類似表記語知識は、たとえば一文字違いの単語の組合せを機械的に列举したり、日本人のつづり誤りのデータを収集するなどによって構築することが可能である。

【0019】この第1実施形態では、「語A」と「語B」とが同一の文書や段落、あるいは文などに出現する場合に、両者は互いに共起語であるということにする。また、「語A」と「語B」とが意味的に関係がある場合に、両者は互いに関連語であるということにする。図3の(知識h)〜(知識l)が共起語知識の例であり、(知識m)〜(知識o)が関連語知識の例である。たとえば、(知識j)から「情報」と「検索」という語は共起することがわかり、(知識m)から「パソコン」と「オフコン」とは意味的に関係の深い語であることがわかる。

【0020】共起語知識は、たとえば既存のかな漢字変換用辞書の共起データを利用して構築することが可能で

ある。また、関連語知識は、たとえば既存のシソーラスにおける兄弟語（同一の親ノードをもつ語）の情報を利用して構築したり、あるいは前述の共起語知識を利用して、「語A」と「語B」とが共起し、かつ「語C」と「語B」とが共起するならば、「語A」と「語C」とは同じような文脈に出現したので関連語である」といった方針により構築することが可能である。なお、類似語知識、共起語知識および関連語知識などの具体的な構築方法は、この発明の主旨ではなく、どのような方法で構築されても構わない。

【0021】図4に、第1実施形態の検索条件入力部11から検索語訂正部12に渡される語入力された検索語を含む検索条件の例を示す。図4(1)～(6)は、複数の検索条件をAND、OR、NOT演算子により結合して検索条件を構成するブール検索の例である。

【0022】図4(1)は、「自然言語」と「解析」という2つの語を共に含む情報を検索しようとしたときに、かな漢字変換を誤って「解析」のかわりに「金庫」と入力してしまった例である。図4(2)は、同様の検索を行うとしたときに、「かいせき」とひらがなを入力した後、漢字に変換しないまま確定してしまった例である。人間がこのような間違いを犯し易いことは、ワープロ文書などを見れば明らかである。

【0023】これらの例の場合、図3の(知識a)を利用すれば、検索語訂正部12には2番目の検索語を「解析」あるいは「懐石」と訂正する必要があるかもしれないということがわかる。

【0024】一方、1番目の検索語である「自然言語」に関する共起語知識としては、図3の(知識h)および(知識i)があるが、このうち(知識i)は「自然言語」と「解析」とが共起することを示しているので、2番目の検索語は「解析」と訂正すればよいことがわかる。

【0025】図4(3)は、NOTを使ったブール検索における語入力の例を示している。「顧客満足度(Customer Satisfaction)」の略語として使われる「CS」という語を含む情報を検索したいユーザが、「CS」には「通信衛星(communication satellites)」の意味もあるので、後者の意味での「CS」という語を含む情報を排除するために、NOT「衛星」と入力しようとして、誤って「衛生」としてしまった場合である。この場合、図3の(知識b)および(知識k)から、「衛生」を「衛星」に訂正すればよいことがわかる。

【0026】図4(1)～(3)は、類義語知識のうちの同音異義語知識を利用したかな漢字変換の誤りを訂正する例であった。これに対し、図4(4)～(5)は、類義語知識のうちの類似表記語知識を利用して入力誤りを訂正する例である。

【0027】図4(4)は、「情報」AND「検索」と

いうブール検索を行うおとして、「検索」のかわりに「検索」と入力してしまった場合である。この場合、図3の(知識f)から、意図した検索語は「検索」と表記の似ている「検索」である可能性があることがわかり、一方、図3の(知識j)から「情報」と「検索」とは共起することがわかるので、「検索」を「検索」に訂正すればよいことがわかる。

【0028】図4(5)は、(4)と同様であるが、検索語が英語である例である。「光学的文字読み取り機(optical character reader)」に関する情報を検索しようとして、「reader」のつづりを間違えて「leader」としてしまった場合である。これは図3の(知識g)および(知識l)により訂正可能である。

【0029】図4(1)～(5)は、類似語知識と共起語知識とを利用して検索語訂正を行なう例であった。これに対し、図4(6)は、類義語知識と関連語知識とを利用して検索語訂正を行なう例である。

【0030】図4(6)は、ANDおよびORを使って「放送衛星」、「衛星放送」、「通信衛星」、「衛星通信」などに関する情報を検索しようとしたユーザが、「放送」のかわりに「包装」と誤入力してしまった場合を示している。この場合、図3の(知識c)から「包装」の同音異義語に「放送」があることがわかり、一方、図3の(知識n)から「放送」の関連語に「通信」があることがわかる。ここで、検索条件では「包装」と「通信」がORで結ばれていたため、「包装」は「放送」に訂正すればよいことがわかる。

【0031】図5に、図2におけるi番目の検索語の誤入力判定アルゴリズム(図2のステップA5)の一例を示す。ここでは、検索条件が図4(1)であり、i番目の検索語が「金庫」であり、かつ検索語訂正用語書16が図3の知識を含む場合を例として説明する。まず、i番目の検索語に関する類似語知識から、i番目の検索語の同音異義語および類似表記語を取出す。図3の(知識a)から「かいせき」、「解析」、「懐石」という同音異義語が得られる(ステップB1)。次に、i番目以外の検索語、すなわち「自然言語」に関する共起語知識/関連語知識のうち、「かいせき」、「解析」、もしくは「懐石」という語を含むエントリを取出す。図3の(知識i) {自然言語、解析} が得られる(ステップB2)。そして、このようなエントリが存在するので(ステップB3のY)、検索語「金庫」は誤入力であり、「解析」に訂正すべきであると判定することができる(ステップB4)。

【0032】図4(1)～(6)の検索条件を本方式により訂正した結果の例を、図6(1)～(6)に示す。図4(1)～(6)は、ブール検索式の例であったが、この発明は検索条件の形式には依存しない。図4(7)～(8)には、ブール検索以外の検索方式を用いた検索

システムにおける検索条件の例が示されている。

【0033】図4(7)は、「自然言語」と「解析」という語が同一段落内に出現するような情報を検索しようとして、誤って「解析」を「会席」としてしまっただけである。一方、図4(8)は、「自然言語」と「解析」という語がともに第1文目に出現するような情報を検索しようとした同様な例である。これらの場合も、図4(1)の場合と同様に訂正が可能である。

【0034】この他、検索語の頻度を指定するような検索や、検索語と検索語との間の距離を指定するような検索などに対しても、同様に検索語訂正を行なうことは有効である。

【0035】(第2実施形態)次に、この発明の第2実施形態について説明する。図7に、第2実施形態に係る情報検索システムの構成を示す。図7に示したように、第1実施形態と第2実施形態とのシステム構成上の違いは、後者が検索結果評価情報入力部18を有する点である。また、システムの処理の流れにおける違いは、前者では入力された検索条件が訂正されてから検索が行なわれるのに対し、後者では入力された検索条件を用いながら検索を行ない、その検索結果に対するユーザの評価情報を得られてはじめて検索条件を訂正する点である。よって、この第2実施形態では、再検索を行なうことを前提としている。以下、第1実施形態と異なる点のみについて詳しく説明する。

【0036】検索部14は、ユーザにより入力された検索条件を用いて第1回目の検索を行ない、検索結果出力部15は、ユーザに検索結果を出力する。次に、ユーザは、この検索結果を見て、自分の要求にあったものであるかどうかの評価を行ない、その評価結果を検索結果評価情報入力部18に入力する。この評価情報は、検索語訂正部12に渡される。検索語訂正部12は、検索結果のうちどの情報がユーザの要求に適合し、どの情報が適合しなかったかを分析し、必要があればこれをもとに検索条件を修正する。そして、検索部14は、この訂正された検索条件を用いて再検索を行なう。

【0037】図8に、第2実施形態の検索結果評価情報入力部18の処理の流れの一例を示す。検索結果評価情報入力部18は、検索結果出力部15によってユーザに出力された情報の各々について、ユーザの評価を受け付ける(ステップC2～ステップC3)。ここで、ユーザの評価とは、検索された情報がユーザにとってどれほど有用であったかという情報を検索システムに与えることを意味する。たとえば「この情報は有用であったが、この情報は有用でなかった」といった2段階の評価や、点数付けによる多段階評価などが考えられる。なお、検索結果に対して評価を行なうこと自体は、たとえば文獻

(「情報検索論」、David Ellis原著、細野公明監訳、丸善)で「適合性判定」として紹介されているように公知の技術であり、この発明の主旨ではない。

そして、検索結果評価情報入力部18は、最後に前述の評価情報を検索語訂正部に渡す(ステップC6)。

【0038】図9に、第2実施形態の検索語訂正部12の処理の流れの一例を示す。検索語訂正部12は、まず検索結果評価情報入力部18から評価情報を受取る(ステップD1)。そして、評価された検索結果の各々について、検索条件がどのようにして満たされたかを分析する(ステップD2)。この分析結果と、検索語訂正用辞書から得た各検索語の類似語知識、共起語知識および関連語知識とを照合して、誤入力されたと思われる検索語を同定しこれを訂正する(ステップD6～ステップD7)。最後に、訂正された検索条件を検索条件記憶部13に渡す(ステップD8)。

【0039】図10に、検索結果評価情報入力部18から検索語訂正部12に渡される評価情報の一例を示す。ここでは、ユーザは、図6(6)の検索条件を入力したつもりであったが、実際には、誤入力のために図4(6)の検索条件が入力されたとする。これにより検索を行なった結果、3つの文書が検索され、これらに対してユーザが2段階の評価を行なったものとする。各文書が検索語を含んでいた場合を1、含んでいなかった場合を0で、また、各文書が検索条件を満たした場合を1、満たさなかった場合を0で示してある。文書1は、たまたま「包装」と「衛星」という語を両方含んでいたために検索条件が満たされていたが、内容的にはユーザの求める「放送衛星」、「衛星放送」、「通信衛星」、「衛星通信」とはまったく無関係であったために、「有用でない」と評価されている。一方、文書2および文書3は、「通信」と「衛星」という語を含むことにより検索条件を満たし、かつユーザに「有用である」と評価されている。

【0040】以下に、図10のような評価結果に対して検索語訂正部12が行なう処理を説明する。図10において、たとえば「有用でない」と評価された文書に出現し、かつ「有用である」と評価された文書に出現しなかった検索語を求めると、「包装」という検索語が得られる。「包装」という語を含んでいた文書1が、検索条件を満たしているにもかかわらず「有用でない」と評価されたことから、「包装」という検索語は誤入力である可能性が大きい。以後、第1実施形態と同様に検索語訂正用辞書16を利用することにより、「包装」を「放送」と訂正することが可能である。

【0041】なお、この第2実施形態におけるユーザの評価情報を利用した検索語訂正処理は、第1実施形態と同様にブル検索以外においても有効であることはいまでもない。

【0042】

【発明の効果】以上詳述したように、この発明によれば、誤入力された語句が自動的に検出・訂正されるため、その作業効率は飛躍的に向上する。また、入力され

た語句が多義に解釈される語句であっても、その解釈を選定することができるため、検索の効率および精度を向上させることが可能となる。さらに、ユーザによる検索結果の評価を語句の検出・訂正にフィードバックすることにより、再検索の精度を向上させることができるようになる。

【図面の簡単な説明】

【図1】この発明の第1実施形態に係る情報検索システムの構成を示す図。

【図2】同実施形態の検索語訂正部の処理の流れの一例を示すフローチャート。

【図3】同実施形態の検索語訂正用辞書に登録されている類似語知識、共起語知識および関連語知識の一例を示す図。

【図4】同実施形態の検索条件入力部から検索語訂正部に渡される語入力された検索語を含む検索条件の例を示す図。

【図5】同実施形態の図2における1番目の検索語の誤

入力判定アルゴリズム（図2のステップA5）の一例を示すフローチャート。

【図6】同実施形態の検索条件を訂正した結果を示す図。

【図7】この発明の第2実施形態に係る情報検索システムの構成を示す図。

【図8】同実施形態の検索結果評価情報入力部の処理の流れの一例を示すフローチャート。

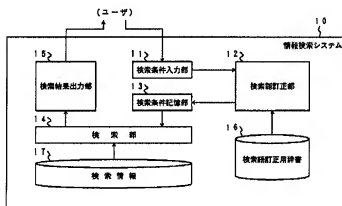
【図9】同実施形態の検索語訂正部の処理の流れの一例を示すフローチャート。

【図10】同実施形態の検索結果評価情報入力部から検索語訂正部に渡される評価情報の一例を示す図。

【符号の説明】

10…情報検索システム、11…検索条件入力部、12…検索語訂正部、13…検索条件記憶部、14…検索部、15…検索結果出力部、16…検索語訂正用辞書、17…検索情報、18…検索結果評価情報入力部。

【図1】



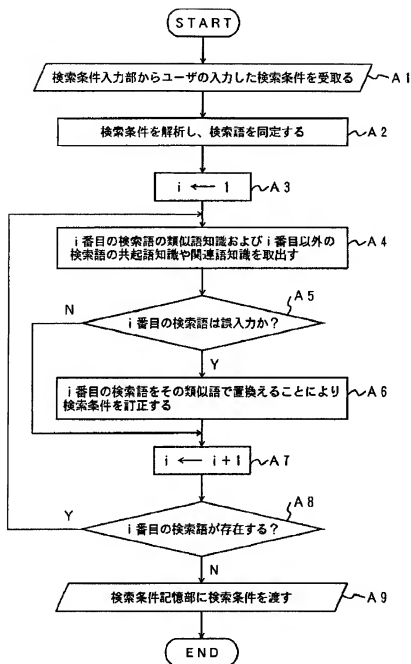
【図4】

(1)	「自然言語」 AND 「金庫」
(2)	「自然言語」 AND 「かいせき」
(3)	「CS」 AND NOT 「衛生」
(4)	「情報」 AND 「数量」
(5)	「optical」 AND 「character」 AND 「reader」
(6)	「 <u>勉強</u> 」 OR 「 <u>通関</u> 」 AND 「 <u>数量</u> 」
(7)	一検索内: 「自然言語」、「金庫」
(8)	1文用: 「自然言語」、「金庫」

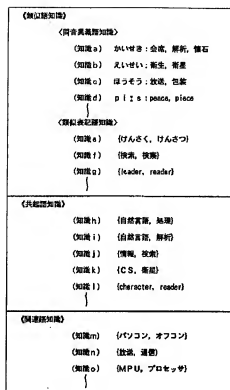
【図6】

(1)	「自然言語」 AND 「 <u>解題</u> 」
(2)	「自然言語」 AND 「 <u>解題</u> 」
(3)	「CS」 AND NOT 「 <u>衛生</u> 」
(4)	「 <u>情報</u> 」 AND 「 <u>検査</u> 」
(5)	「optical」 AND 「character」 AND 「 <u>reader</u> 」
(6)	「 <u>勉強</u> 」 OR 「 <u>通関</u> 」 AND 「 <u>数量</u> 」

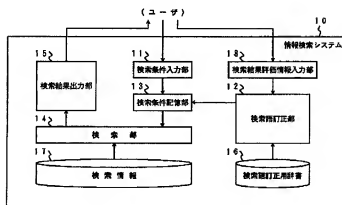
【図2】



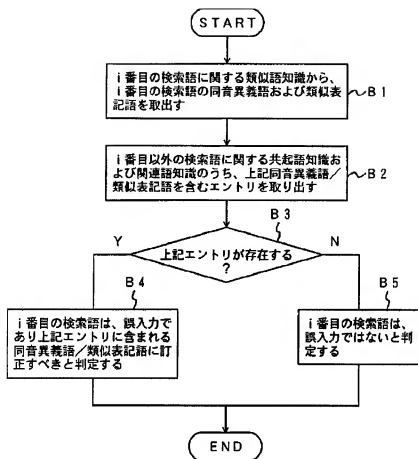
【図3】



【図7】



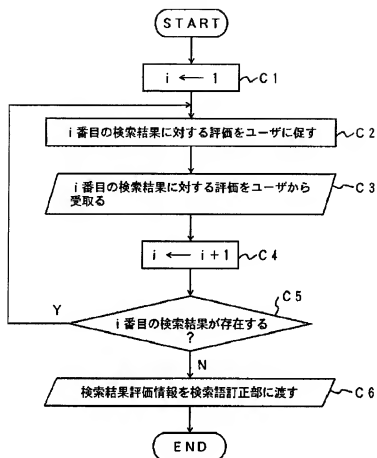
【図5】



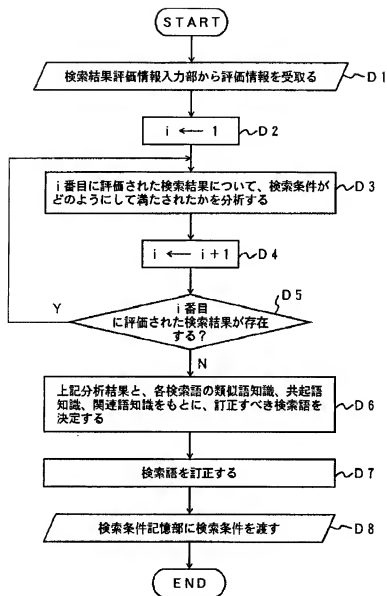
【図10】

	包接	漏脱	衝突	(包接 OR 漏脱) AND 衝突	評価結果
文 書 1	1	0	1	1	有用でない
文 書 2	0	1	1	1	有用である
文 書 3	0	1	1	1	有用である

【図8】



【図9】



H10-240739

Machine Translation from Japanese Patent & Utility Model Gazette DB
(<http://www4.ipdl.inpit.go.jp/Tokujitu/tjsogodben.ipdl?N0000=115>)

CLAIMS

[Claim(s)]

[Claim 1]An information retrieval device with which information corresponding to a search condition constituted including words and phrases inputted as a keyword is retrieved, comprising:

A dictionary for words-and-phrases correction containing at least one or more of similar-words knowledge, coincidence word knowledge, and the related term knowledge.

An erroneous input words-and-phrases detection means by which said dictionary for search term correction detects words and phrases by which the erroneous input was carried out.

A words-and-phrases correcting means which corrects words and phrases which said erroneous input words-and-phrases detection means detected in said dictionary for search term correction.

[Claim 2]An information retrieval device with which information corresponding to a search condition constituted including words and phrases inputted as a keyword is retrieved, comprising:

A dictionary for words-and-phrases selection containing at least one or more of similar-words knowledge, coincidence word knowledge, and the related term knowledge.

A words-and-phrases selection means selected to one of interpretations in said dictionary for words-and-phrases selection when said inputted words and phrases are words and phrases interpreted by polysemy.

[Claim 3]The information retrieval device according to claim 1 or 2 which possesses further an evaluation input means to input evaluation which shows the degree of usefulness of each retrieved information, and a feedback means which analyzes evaluation inputted by said evaluation input means, and corrects said words and phrases, and is characterized by things.

[Claim 4]In an information retrieval method which retrieves information corresponding to a search condition constituted including words and phrases which were provided with a dictionary for words-and-phrases correction containing at least one or more of similar-words knowledge, coincidence word knowledge, and the related term knowledge, and were inputted as a keyword, An information retrieval method performing a search using a search condition which detects words and phrases by which the erroneous input was carried out in said dictionary for search term correction, corrects this detected words-and-phrases word in said dictionary for words-and-phrases correction, and is built including words and phrases after this correction.

[Claim 5]In an information retrieval method which retrieves information corresponding to a search condition constituted including words and phrases which were provided with a dictionary for words-and-phrases selection containing at least one or more of similar-words knowledge, coincidence word knowledge, and the related term knowledge, and

were inputted as a keyword, An information retrieval method performing a search after selecting to one of interpretations in said dictionary for words-and-phrases selection and making this selection result reflect in said search condition, when said inputted words and phrases are words and phrases interpreted by polysemy.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001] **[Field of the Invention]** This invention relates to the information retrieval device and information retrieval method which detect and correct automatically the words and phrases by which started the information retrieval device and information retrieval method which retrieve the information corresponding to the search condition constituted including the words and phrases inputted as a keyword, especially the erroneous input was carried out, and perform a search.

[0002] **[Description of the Prior Art]** In recent years, the media of information have also diversified [which is increasing steadily] the quantity of the information which an individual can access a text, a picture, a sound, etc. with the spread of a personal computer, the Internet, Electronic Libraries, etc. And the demand of the advanced information retrieval systems with which only the information searched for out of huge information by such a situation is retrieved is increasing.

[0003] When using the usual information retrieval system, a user constitutes the search condition of the form that he inputs one or more search terms, and can understand a system, in order to tell a demand of him to a system. A system is a basis of the assumption "a user's demand = search condition", and elects and outputs only the information by which a search condition is fulfilled out of a retrieval object.

[0004] However, the assumption "a user's demand = search condition" is not realized in practice in many cases. When specifying a search condition via input devices, such as a keyboard, a character reader, voice recognition equipment, especially, the search condition which was widely different with a user's intention for the erroneous input of a search term may be created. For example, when it is going to specify the search term of a Chinese character using a kana-kanji conversion system, the homonym which is completely unrelated may be inputted by a conversion error.

[0005] Although detection and correction of an erroneous input are possible by performing word dictionary length to all the search terms to the word etc. of the spelling which is not possible, Like the homonymic example mentioned above, when right words and phrases had been obtained as a word as a result of a word input, with the conventional information retrieval system, were able to detect this and it was not able to be corrected. For this reason, search results mistaken according to the mistaken search condition were obtained, and the case where a user noticed his erroneous input for the first time there had arisen. Even if search results mistaken by the erroneous input of the search term were obtained, in addition, there was also a case where a user did not notice an erroneous input by himself.

[0006] **[Problem(s) to be Solved by the Invention]** Thus, in the conventional information retrieval system, when the erroneous input of the right words and phrases had been carried out as a search term as a word, there was a problem that could detect this and it could not be corrected.

[0007]This invention is made in view of such the actual condition, and is a thing. The purpose is to provide the information retrieval device and information retrieval method which detect and correct ***** automatically and perform a search.

[0008] **[Means for Solving the Problem]**A dictionary including related term knowledge for obtaining coincidence word knowledge and a related term for acquiring similar-words knowledge for obtaining similar words of words and phrases inputted, for example as a keyword in order that this invention may attain the purpose mentioned above, and a coincidence word, etc. is used, After detecting and correcting words and phrases by which the erroneous input was carried out, it is made to perform a search.

[0009]Efficiency and accuracy of search are raised by selecting an interpretation of words and phrases interpreted by polysemy using this dictionary. Accuracy of re retrieval is raised by analyzing evaluation of search results by a user and feeding back this analysis result to detection and correction of words and phrases by which the erroneous input was carried out.

[0010]Thus, even if it is a case where an erroneous input occurs, when a user creates a search condition intended originally and performs a search, it makes it possible to provide suitable information retrieval environment.

[0011] **[Embodiment of the Invention]** Hereafter, this embodiment of the invention is described with reference to Drawings.

(A 1st embodiment) A 1st embodiment of this invention is described first. The composition of the information retrieval system concerning a 1st embodiment is shown in drawing 1. As shown in drawing 1, this information retrieval system 10 consists of the search condition input section 11, the search term correction part 12, the search condition storage parts store 13, the retrieval part 14, and the search-results outputting part 15. Here the search condition input section 11 to input devices, such as a keyboard, a character reader, and voice recognition equipment, the search-results outputting part 15, The search condition storage parts store 13 corresponds to output units, such as a display and a printer, at main memory, a hard disk drive, etc., respectively, and the retrieval part 14 and the search term correction part 12 correspond to the program in which execution control is carried out by CPU.

[0012]The search condition inputted into the search condition input section 11 by the user is passed to the search term correction part 12, and correction of the words and phrases which serve as a keyword if needed, i.e., a search term, is performed. The corrected search condition is memorized by the search condition storage parts store 13, and the retrieval part 14 retrieves information according to this search condition. Search results are outputted to a user by the search-results outputting part 15.

[0013]The difference between the conventional information retrieval system and this information retrieval system 10 is only the point that a search condition is once passed in the latter to the search term correction part 12, and the search condition storage parts store 13 is passed after that to the search condition inputted into the search condition input section 11 being passed to the direct retrieval condition storage section 13, in the former. Therefore, a search system like the existing throat may be sufficient except processing of the search term correction part 12. If it searches when a user specifies one or more search terms, the specification method of a search condition, a retrieval object, a search method, etc. will not be asked. What kind of language [besides Japanese and

English] may be sufficient as a search term? Hereafter, it explains focusing on operation of the search term correction part 12.

[0014]An example of the flow of processing of the search term correction part 12 of a 1st embodiment is shown in drawing 2. The search term correction part 12 receives first the search condition which the user inputted from the search condition input section 11 (Step A1), analyzes this, and identifies a search term (Step A2). And the following processings are performed about each search term.

[0015]The similar-words knowledge of a search term, and the coincidence word knowledge and related term knowledge of the other search term to which its attention is paid now are taken out from the dictionary for search term correction (step A4). Next, it is judged using these knowledge whether the search term to which its attention is paid now is an erroneous input (step A5). When it judges with it being an erroneous input, a search condition is corrected by replacing (Y of step A5), and a search term with the above-mentioned similar words (Step A6).

[0016]Only detection of an erroneous input is performed first, the message of "whether to correct" is displayed on a user, and it may be made to correct instead of correcting a search term with a perfect automatic interactively.

[0017]An example of the similar-words knowledge registered into the dictionary 16 for search term correction of a 1st embodiment, coincidence word knowledge, and related term knowledge is shown in drawing 3. According to this 1st embodiment, when "the word A" turns into "the word B" by erroneous inputs, such as a kana-kanji conversion error and a spelling error, it uses that "the word A" and the "word B" are similar words mutually.

[0018]Two kinds such as homonym knowledge and similar notation word knowledge are shown as similar-words knowledge by the example of drawing 3. - (knowledge a) (knowledge d) of drawing 3 is an example of homonym knowledge, and - (knowledge e) (knowledge g) is an example of similar notation word knowledge. For example, (the knowledge a) shows that both the words of a "place of meeting" and "analysis" have reading of a "paddle cough", and it turns out that the notation is alike and it is easy to carry out the erroneous input of the word of "leader" and "reader" from (the knowledge g). This homonym knowledge can be built, for example using the existing dictionary for kana-kanji conversions. Similar notation knowledge can be built by enumerating the combination of the word of a single-character difference mechanically, for example, or Japanese people spelling and collecting the data of an error etc.

[0019]According to this 1st embodiment, when "the word A" and the "word B" appear in the same document and paragraph, or a sentence, both use that it is a coincidence word mutually. When "the word A" and the "word B" are semantically related, both use that it is a related term mutually. - (knowledge h) (knowledge l) of drawing 3 is an example of coincidence word knowledge, and - (knowledge m) (knowledge o) is an example of related term knowledge. For example, as for "information" and the word of "search", (the knowledge j) shows coinciding in many cases, and (the knowledge m) shows that a "personal computer" and an "office computer" are closely related words semantically.

[0020]Coincidence word knowledge can be built, for example using the coincidence data of the existing dictionary for kana-kanji conversions. The coincidence word knowledge of the above-mentioned [**** / building related term knowledge, for example using the information on the brother language (word with the same parent node) in the existing

thesaurus] is used, If "the word B" coincides with ""word A" and "the word C" and the "word B" coincide, it is possible to build by "the word A" and the plan "it to be a related term since word C" appeared in the same context." Concrete constructing methods, such as similar-words knowledge, coincidence word knowledge, and related term knowledge, may be built by the not main point but what kind of method of this invention.

[0021]The example of the search condition containing the search term which is passed to the search term correction part 12 from the search condition input section 11 of a 1st embodiment, and by which the word input was carried out is shown in drawing 4.

Drawing 4 (1) - (6) is an example of the Boolean search which combines two or more search terms by AND, OR, and a NOT-operation child, and constitutes a search condition.

[0022]Drawing 4 (1) is an example which has inputted the kana-kanji conversion as the "place of meeting" instead of "analysis" accidentally, when it is going to retrieve information including both "natural language" and two words of "analysis." When drawing 4's (2)'s trying to perform same search, after inputting a "paddle cough" and a hiragana, it is an example become final and conclusive not changed into a Chinese character. It is clear that human being tends to make such a mistake, if a word processor document etc. are seen.

[0023]If the (knowledge a) of drawing 3 is used in the case of these examples, it turns out that it may be necessary to correct the 2nd search term with "analysis" or a "tea lunch" at the search term correction part 12.

[0024]On the other hand, although there are the (knowledge h) and the (knowledge i) of drawing 3 as coincidence word knowledge about the "natural language" which is the 1st search term, among these since it is shown that "natural language" and "analysis" coincide, as for the (knowledge i), it turns out that the 2nd search term should just correct with "analysis."

[0025]Drawing 4 (3) shows the example of the erroneous input in the Boolean search using NOT. The user who wants to retrieve information including the word of "CS" used as an abbreviation of "customer satisfaction (Customer Satisfaction)", Since "CS" has a meaning of "a communications satellite (communication satellites)", in order to eliminate information including the word of "CS" in the latter meaning, it is a case where tended to input NOT "satellite" and it keeps as "health" accidentally. In this case, the (knowledge b) and the (knowledge k) of drawing 3 show that what is necessary is just to correct "health" to a "satellite."

[0026]Drawing 4 (1) - (3) was an example which corrects the error of a kana-kanji conversion using the homonym knowledge of the synonym knowledge. On the other hand, drawing 4 (4) - (5) is an example which corrects an input error using the similar notation word knowledge of the synonym knowledge.

[0027]Drawing 4 (4) is a case where tried to perform Boolean search called "information" AND "search", and "criminal investigation" has been inputted instead of "search." In this case, since the (knowledge f) of drawing 3 shows that the intended search term may be "search" to which the notation resembles "criminal investigation" and, as for "information" and "search", the (knowledge j) of drawing 3 shows coinciding on the other hand, it turns out that what is necessary is just to correct "criminal investigation" to "search."

[0028]Although drawing 4 (5) is the same as that of (4), a search term is an example

which is English. It is a case where tended to retrieve the information about "an optical-character-recognition machine (optical character reader)", made a mistake in spelling of "reader", and it keeps as "leader." This can be corrected by the (knowledge g) and the (knowledge l) of drawing 3.

[0029]Drawing 4 (1) - (5) was an example which performs search term correction using similar-words knowledge and coincidence word knowledge. On the other hand, drawing 4 (6) is an example which performs search term correction using synonym knowledge and related term knowledge.

[0030]Drawing 4 (6) shows the case where the user who tried to retrieve the information about a "broadcasting satellite", "satellite broadcasting", a "communications satellite", "satellite communication", etc. using AND and OR has done the erroneous input to the "package" instead of "broadcast." In this case, the (knowledge c) of drawing 3 shows that the homonym of a "package" has "broadcast", and, on the other hand, it turns out that the related term of "broadcast" has "communication" from the (knowledge n) of drawing 3. Here, since a "package" and "communication" were connected with the search condition by OR, it turns out that what is necessary is just to correct a "package" to "broadcast."

[0031]An example of the erroneous input judging algorithm (step A5 of drawing 2) of the i-th search term in drawing 2 is shown in drawing 5. Here, a search condition is drawing 4 (1), and the i-th search term is a "place of meeting", and the case where the dictionary 16 for search term correction includes the knowledge of drawing 3 is explained as an example. First, if the homonym and similar notation word of the i-th search term are taken out from the similar-words knowledge about the i-th search term, the homonym a "paddle cough", "analysis", and a "tea lunch" will be obtained from the (knowledge a) of drawing 3 (Step B1). Next, if a "paddle cough", "analysis", or an entry including the word of a "tea lunch" is taken out among the coincidence word knowledge / related term knowledge about search terms other than the i-th, i.e., "natural language", {(knowledge i) the natural language and analysis} of drawing 3 will be obtained (step B-2). And since such an entry exists (Y of Step B3), it can judge with the search term "place of meeting" being an erroneous input, and being corrected to "analysis" (step B4).

[0032]Drawing 4 (1) The example of the result of having corrected the search condition of - (6) with this method is shown in drawing 6 (1) - (6). Drawing 4 (1) Although - (6) was an example of the Boolean search type, it does not depend for this invention on the form of a search condition. Drawing 4 (7) The example of the search condition in the search system which used retrieval systems other than Boolean search is shown in - (8).

[0033]Drawing 4 (7) is the example which was going to retrieve the information that "natural language" and the word of "analysis" appeared in the same paragraph, and kept "analysis" as a "place of meeting" accidentally. On the other hand, drawing 4 (8) is the same example that was going to retrieve the information that both "natural language" and the word of "analysis" appeared in the 1st sentence. It can correct like the case of drawing 4 (1) also in these cases.

[0034]In addition, it is effective to perform search term correction similarly to search which specifies the frequency of a search term, search which specifies the distance between search terms, etc.

[0035](A 2nd embodiment), next a 2nd embodiment of this invention are described. The composition of the information retrieval system concerning a 2nd embodiment is shown in drawing 7. As shown in drawing 7, the difference on the system configuration of a 1st

embodiment and a 2nd embodiment is the point that the latter has the search-results evaluation information input part 18. The difference in the flow of processing of a system is the point of correcting a search condition only after it refers to the former first using the search condition inputted by the latter to search being performed after the inputted search condition is corrected, and a user's evaluation information over the search results is acquired. Therefore, it is premised on performing re retrieval in this 2nd embodiment. Hereafter, only a different point from a 1st embodiment is explained in detail.

[0036]The retrieval part 14 performs 1st search using the search condition inputted by the user, and the search-results outputting part 15 outputs search results to a user. Next, a user evaluates whether it used to be in a demand of it, seeing these search results, and inputs that evaluation result into the search-results evaluation information input part 18. This evaluation information is passed to the search term correction part 12. The search term correction part 12 analyzes which information which information suited a user's demand among search results, and did not suit, and it amends a search condition based on this if necessary. And the retrieval part 14 performs re retrieval using this corrected search condition.

[0037]An example of the flow of processing of the search-results evaluation information input part 18 of a 2nd embodiment is shown in drawing 8. The search-results evaluation information input part 18 receives a user's evaluation about each of the information outputted to the user by the search-results outputting part 15 (Step C2 - Step C3). Here, a user's evaluation means giving the information whether the retrieved information was useful however for the user to a search system. For example, two steps of evaluations "this information was not useful although this information was useful", the multi stage story evaluation by mark attachment, etc. can be considered. Evaluating to search results itself is publicly known art as introduced, for example as "a conformity judging" by document (an "information retrieval theory", David Ellis original paper, Kimio Hosono supervision of translation, Maruzen), and it is not the main point of this invention. And the search-results evaluation information input part 18 passes the above-mentioned evaluation information to the last at a search term correction part (Step C6).

[0038]An example of the flow of processing of the search term correction part 12 of a 2nd embodiment is shown in drawing 9. The search term correction part 12 receives evaluation information from the search-results evaluation information input part 18 first (Step D1). And it is analyzed how the search condition was fulfilled about each of the evaluated search results (Step D2). The search term which compares this analysis result, and the similar-words knowledge, coincidence word knowledge and related term knowledge of each search term obtained from the dictionary for search term correction, and is considered that the erroneous input was carried out is identified, and this is corrected (Step D6 - Step D7). At the end, the corrected search condition is passed to the search condition storage parts store 13 (Step D8).

[0039]An example of the evaluation information passed to the search term correction part 12 from the search-results evaluation information input part 18 is shown in drawing 10. Here, although the user thought that he inputted the search condition of drawing 6 (6), he presupposes actually that the search condition of drawing 4 (6) was inputted for the erroneous input. As a result of this searching, three documents should be searched and the user should perform two-step evaluation to these. Are 0 about 1 and the case where it does not contain, in the case where each document contains the search term, and 1 shows

the case where each document fulfills a search condition, and 0 has shown the case where it does not fill. Since the document 1 included both of words of a "package" and a "satellite" by chance, the search condition has been fulfilled, but since it was completely unrelated, the "broadcasting satellite", "satellite broadcasting", the "communications satellite", and the "satellite communication" for which a user asks in contents are estimated not to be "useful." On the other hand, by including the word of "communication" and a "satellite", the document 2 and the document 3 fulfill a search condition, and are estimated to be "useful" by the user.

[0040]The processing which the search term correction part 12 performs to below to an evaluation result like drawing 10 is explained. In drawing 10, if it asks for the search term which did not appear in the document which appeared in the document estimated not to be "useful", and was estimated to be "useful", the search term a "package" will be obtained. Although the document 1 included the word of a "package" fulfilled the search condition, since it was estimated not to be "useful", the search term a "package" may be an erroneous input. Henceforth, it is possible by using the dictionary 16 for search term correction like a 1st embodiment to correct a "package" with "broadcast."

[0041]It cannot be overemphasized that the search term correction processing using the user's evaluation information in this 2nd embodiment is effective like a 1st embodiment in addition to Boolean search.

[0042]

[Effect of the Invention]Since the words and phrases by which the erroneous input was carried out are automatically detected and corrected according to this invention as explained in full detail above, that working efficiency improves by leaps and bounds. Since the interpretation can be selected even if the inputted words and phrases are words and phrases interpreted by polysemy, it becomes possible to raise the efficiency and accuracy of search. The accuracy of re retrieval can be raised now by feeding back evaluation of the search results by a user to detection and correction of words and phrases.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1]The figure showing the composition of the information retrieval system concerning a 1st embodiment of this invention.

[Drawing 2]The flow chart which shows an example of the flow of processing of the search term correction part of the embodiment.

[Drawing 3]The figure showing an example of the similar-words knowledge registered into the dictionary for search term correction of the embodiment, coincidence word knowledge, and related term knowledge.

[Drawing 4]The figure showing the example of the search condition containing the search term which is passed to a search term correction part from the search condition input section of the embodiment, and by which the word input was carried out.

[Drawing 5]The flow chart which shows an example of the erroneous input judging algorithm (step A5 of drawing 2) of the i-th search term in drawing 2 of the embodiment.

[Drawing 6]The figure which illustrates the result of having corrected the search condition of the embodiment.

[Drawing 7]The figure showing the composition of the information retrieval system concerning a 2nd embodiment of this invention.

[Drawing 8]The flow chart which shows an example of the flow of processing of the search-results evaluation information input part of the embodiment.

[Drawing 9]The flow chart which shows an example of the flow of processing of the search term correction part of the embodiment.

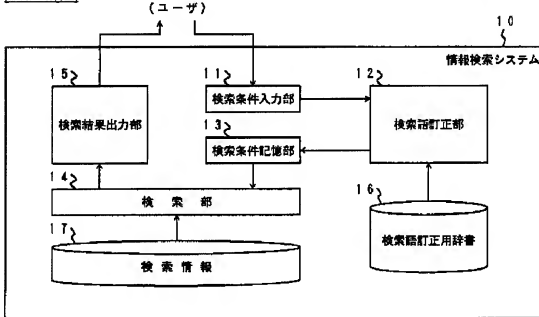
[Drawing 10]The figure showing an example of the evaluation information passed to a search term correction part from the search-results evaluation information input part of the embodiment.

[Description of Notations]

10 [-- A search condition storage parts store, 14 / -- A retrieval part, 15 / -- A search-results outputting part, 16 / -- The dictionary for search term correction, 17 / -- Search information, 18 / -- Search-results evaluation information input part.] -- An information retrieval system, 11 -- A search condition input section, 12 -- A search term correction part, 13

DRAWINGS

[Drawing 1]



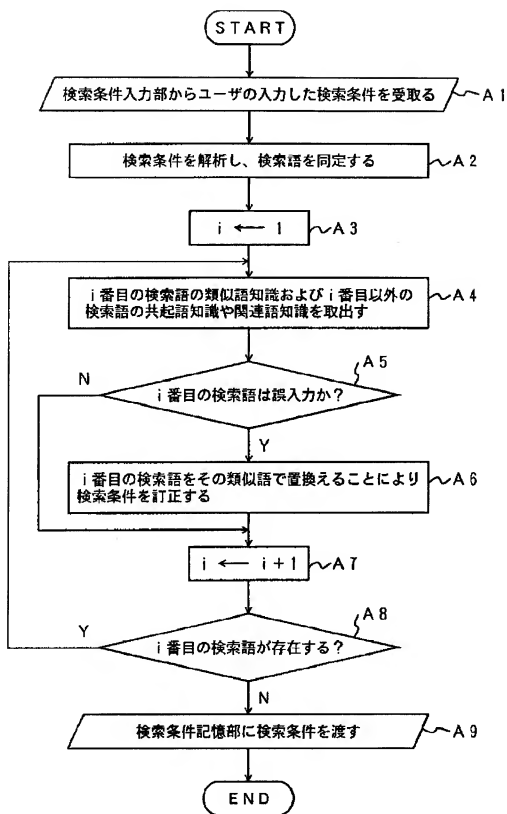
[Drawing 4]

- (1) 「自然言語」 AND 「会席」
- (2) 「自然言語」 AND 「かいせき」
- (3) 「C S」 AND NOT 「衛生」
- (4) 「情報」 AND 「検索」
- (5) 「optical」 AND 「character」 AND 「reader」
- (6) (「包装」 OR 「通信」) AND 「衛星」
- (7) 一段落内：「自然言語」, 「会席」
- (8) 1文目：「自然言語」, 「会席」

[Drawing 6]

- (1) 「自然言語」 AND 「解析」
- (2) 「自然言語」 AND 「解析」
- (3) 「C S」 AND NOT 「衛星」
- (4) 「情報」 AND 「検索」
- (5) 「optical」 AND 「character」 AND 「reader」
- (6) (「放送」 OR 「通信」) AND 「衛星」

[Drawing 2]



[Drawing 3]

《類似語知識》

〈同音異義語知識〉

(知識a) かいせき: 会席, 解析, 懐石

(知識b) えいせい: 衛生, 衛星

(知識c) ほうそう: 放送, 包装

(知識d) p i : s : peace, piece

}

〈類似表記語知識〉

(知識e) {けんさく, けんさつ}

(知識f) {検索, 検索}

(知識g) {leader, reader}

}

《共起語知識》

(知識h) {自然言語, 処理}

(知識i) {自然言語, 解析}

(知識j) {情報, 検索}

(知識k) {CS, 衛星}

(知識l) {character, reader}

}

《関連語知識》

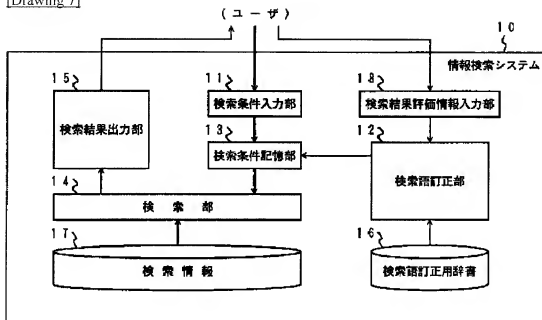
(知識m) {パソコン, オフコン}

(知識n) {放送, 通信}

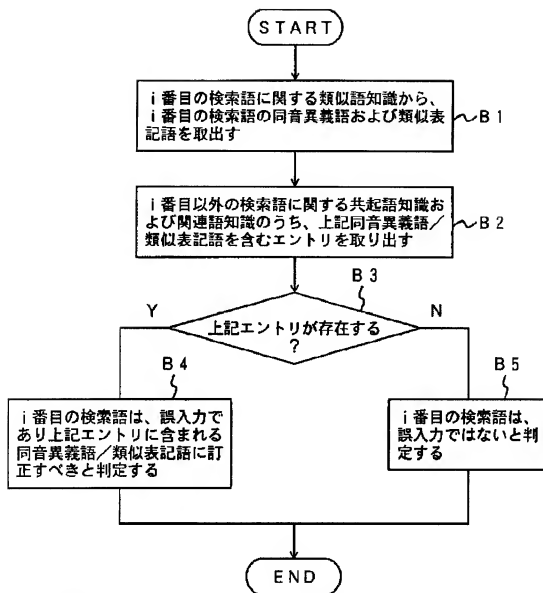
(知識o) {MPU, プロセッサ}

}

[Drawing 7]



[Drawing 5]



[Drawing 10]

	包装	通信	衛星	(包装 OR 通信) AND 衛星	評価結果
文 書 1	1	0	1	1	有用でない
文 書 2	0	1	1	1	有用である
文 書 3	0	1	1	1	有用である

[Drawing 8]

